系統的項目分類の一方法

水　野　欽　司

1. 項目の分類

一般に社会現象，人間現象について分析しようとするとき，まずはじめに行なうことは現象を構成するさまざまな側面についてデータを収集することである。次いで各側面（項目あるいは変数と呼ばれる）の相互類似関係から全体を効率よく要約して現象探究の緒口をつかもうとする。

データが数値として観察単位ごとに得られ，全体として行列の形をなしているとき，この目的に役立つのが因子分析，成分分析である。これらの方法は別のねらいのために使われる場合ももちろん多い。しかし分析の初期段階で，内容の似かよった項目あるいは変数どうしを「分類」し，些細な変異を除去して全体の展望を得るために使われる例が多い。

因子分析等を用いる項目分類では，因子抽出の操作によって全項目をたかだか数個の基本的な座標軸上の点として表わし，それによって分類を行なう。そこではその他の軸での小さい差異は無視して行なうことになる。しかし，これを実際に行なってみると，決して容易な作業ではない。

本稿で扱うのは，このような項目の分類を因子分析などによらずに直接行なおうとする試みである。

すなわち機械的な手順により，最終的に全項目をその親疎の関係に照していずれかのグループに帰せしめようとするものであり，単純なくりかえしを迅速に行う電子計算機の特性を利用した自動分類である。このように，電子計算機の高速性を活用して多くのデータを分類処理しようとする試みは，一般にクラスター分析（cluster analysis）と呼ばれているが，ここで述べる内容もその系列に属するといえる。

本稿における項目分類の基本原理は，項目グループの重心からの距離の2乗平均Dを考え，これを極力小さくするように，各項目をそれぞれ近いグループに配分することである。項目分類は現象分析の初期における粗い見通しを得るためであるが，一方‘類似項目を集める’，という操作は心理テストを作成する際の合成変量（合成得点）を作る手続であることを考慮し，上の測度Dを用いた。群セントロイド法（Group centroid method）など因子分析的な方法に簡単に移行できると同時に，‘グループ’の意味を明確にすることができる。多くの項目を少数のグループにまとめあげる仕方としては，系統クラスター化の方法を採用する。これは最大の分割状態（各項目がそれぞれ1グループをなす）から出発して系統樹的に逐次合併を続けて所定のグループ数が得られたら打切るという方法である。これは計算量を減らすという実用的なねらいによっている。

本稿では，まず，方法の要点と因子分析などとの関連について述べ，次に具体的な計算法と計算例を挙げる。

2. 項目の類似性の指標—集中度D

以下で述べる項目分類で，中心的役割を果すのは項目相互の類似性の指標Dである。まず，‘項目グループの重心からの距離の2乗平均’と定義するこの測度Dが数量的にどう表現されるか，また，いかなる意味をもつか，をみよう。

対象 i に対する項目 j の観測値x_{ij}が下図のようなデータ行列の形で与えられているとする。「項目（あるいは変数)」は，心理テストにおける個々の測定項目，「対象」はテストを受ける個人であるような場合を考えればよい。一般に対象の数Nは項目の数 n よりも大きいとしておく。

対象＼項目	1	2 …… j …… n				平均
1	x_{11}	x_{12}	…… x_{1j}	……	x_{1n}	$x_{1\cdot}$
2	x_{21}	x_{22}	…… x_{2j}	……	x_{2n}	
⋮						
i	x_{i1}	x_{i2}	…… x_{ij}	……	x_{in}	$x_{i\cdot}$
⋮						
N	x_{N1}	x_{N2}	…… x_{Nj}	……	x_{Nn}	$x_{N\cdot}$
平均	$x_{\cdot 1}$	$x_{\cdot 2}$	…… $x_{\cdot j}$	……	$x_{\cdot n}$	$x_{\cdot\cdot}$

データ行列

上図のデータ行列より，n 個の項目をN次元空間の n 本のベクトルとみることができる（図1）。

このn本のベクトルの終点によって各項目を表わすと

し，これらのベクトルの終点に関するn次元を越えない空間内における重心を考え，各終点が重心のまわりでもつ距離の2乗平均を測度Dとする。測度Dは，以上から明らかなように項目をベクトルとみなすとき，その終点

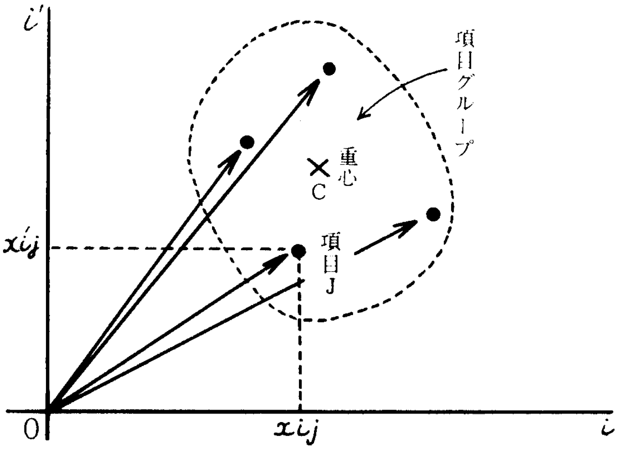

図1　項目の空間表示

の‘集中度’を意味する。この値が小さいほど集中の度合が大きい。言いかえれば，項目相互の類似性（各対象の項目に対する反応模様が似ているという意味での）が高いといえる。

項目間の同質性をベクトル終点のちらばりとして直接的に表現する仕方は，京極[2)]がすでに多元度数分布の数量化に際して，‘重心との距離の平方和’として行なっており，ここでいうDも本質的に同じである。ただし，ここではいくつかの項目グループに関して集中度を比較するという立場をとるので，グループに所属する項目の数のちがいを均等化し，空間上のひろがりの範囲の大小を反映するように2乗の平均を用いている。

測度Dは次のように表現することができる。まず重心Cの座標を，

$$(x_{1\cdot},\ x_{2\cdot},\ \cdots\cdots\cdots\cdots,\ x_{i\cdot},\ \cdots\cdots\cdots\cdots x_{N\cdot})$$

とし，各項目jと重心との距離をd_{jc}とすると，

$$D=\sum_j d_{jc}^2/n=\sum_i\sum_j(x_{ij}-x_{i\cdot})^2/n$$
$$=\frac{1}{n}\sum_i\sum_j x_{ij}^2-\sum_i x_{i\cdot}^2$$

ここで$x_{i\cdot}$ $(i=1,2,\cdots,N)$ は，

$$x_{i\cdot}=\sum_j x_{ij}/n$$

であるから，

$$D=\frac{1}{n}\sum_i\sum_j x_{ij}^2-\frac{1}{n^2}\sum_i\sum_j\sum_k x_{ij}x_{ik}$$
$$(j,k=1,2,\cdots,\ n;\ i=1,2,\cdots,N)$$

ここで各項目の分散を σ_j^2，項目間共分散を σ_{jk} で表わすことにすると，D は，

$$\sigma_j^{\ 2}=\sum_i(x_{ij}-x_{\cdot j})^2/N$$

$$\sigma_{jk}=\sum_i(x_{ij}-\bar{x}_{\cdot j})(x_{ik}-x_{\cdot k})/N$$

$$D=N\left\{\sum_j(\sigma_j^{\ 2}+x_{\cdot j}^{\ 2})/n-\sum_j\sum_k(\sigma_{jk}+x_{\cdot j}\,x_{\cdot k})/n^2\right\}$$

右辺｛　｝の中の第二項は原点と重心との距離の2乗に当る。Nはスケール因子に過ぎないので，以下これを除いて扱うことにすると，

$$D=\left\{\sum_j\sigma_j^{\ 2}/n-\sum_j\sum_k\sigma_{jk}/n^2\right\}+\sum_j(x_{\cdot j}-\bar{x}_{\cdot\cdot})^2/n \quad (1)$$

以上は，項目別の平均，分散になにも制限を付けない場合である。データ処理の実際場面では，個々の項目の平均値の水準の高低に絶対的な意味を与えられない場合，さらに測定の単位そのものに意味がない場合が多い。データ行列がその意味において基準化されているとすれば，(2)，(3)の式のように簡単となる。

項目平均 $\bar{x}_{\cdot j}$ $(j=1,2,\cdots,n)$ が0であれば，（空間の原点を平均にとれば）

$$\overset{(s)}{D}=\sum_j\sigma_j^{\ 2}/n-\sum_j\sum_k\sigma_{jk}/n^2\cdots\cdots\cdots\cdots(2)$$

さらに，各分散 σ_j^2 が1であれば，項目jとkの相関を r_{jk} として，

$$\overset{(r)}{D}=1-\sum_j\sum_k r_{jk}/n^2\cdots\cdots\cdots\cdots(3)$$

以上のように D は，データ行列そのものによらずとも，項目相互間の $(n\times n)$ の内積行列によって表わすことができる。

なお，データが平均 0，分散 1 に基準化されているときは，上で述べた空間は，半径 1 の n 次元の球体に当る。これはいわゆる因子（因子負荷量）空間であり，重心と原点を結ぶ軸はセントロイド第1因子軸であり，$D^{(r)}$ はこの軸上における球表面と重心との距離の2乗である。

3.　項目グループとその合成得点

いくつかの項目が同じグループに所属するということは，なにを意味するか。

いま項目 $X_1,X_2,\cdots\cdots\cdots,X_n$ が同じグループに属するとする。このとき，DのN倍は

$$ND=\sum_i\left\{\sum_j x_{ij}^{\ 2}/n-x_{i\cdot}^{\ 2}\right\}$$

であることから明らかなように，データ行列の「行」の分散の和である。したがって D が小さいことは，対象（行）におけるそれぞれの項目測定値がほぼ等しい——項目相互が同質的な内容を含んでいることを表わす。

これら項目を単純に加算して平均をとり，代表値 Y

を作る。

$$Y=(X_1+X_2+\cdots\cdots\cdots+X_n)/n$$

この Y の平均 $\bar{Y}$ と分散 $\sigma_Y{}^2$ は，

$$\begin{cases} \bar{Y}=\sum_j \bar{x}_{\cdot j}/n \\ \sigma_Y{}^2=\sum_j\sum_k \sigma_{jk}/n^2=\sum_j \sigma_j{}^2/n+\sum_j(\bar{x}_{\cdot j}-\bar{x}_{\cdot\cdot})^2/n-D \end{cases}$$

この場合，各項目が平均 0 とすれば，

$$\sigma_Y{}^2=\sum_j \sigma_j{}^2/n-D^{(s)}$$

また分散も 1 に基準化されているとすれば，

$$\sigma_Y{}^2=1-D^{(r)}$$

である。$D^{(r)}$，$D^{(s)}$ はこれら項目の合成変量の分散を決めている。

Yと個々の項目 X_j との相関 r (Y, X_j) は，

$$r(Y,X_j)=\sum_k \sigma_{jk}/(n\sigma_j\sigma_Y)=\sum_k \sigma_{jk}/\sigma_j\sqrt{\sum_j\sum_k \sigma_{jk}}$$

X_j が平均 0，分散 1 に基準化されているとすれば，

$$r(Y,X_j)=\sum_k r_{jk}/\sqrt{\sum_j\sum_k r_{jk}}$$

この場合，グループの重心を通る仮想項目 Z を考えれば，Z と個別項目 X_j との相関は Z をセントロイド第1因子とする X_j の因子負荷量である。

$$r(Z,X_j)=\sum_k r_{jk}/\sqrt{\sum_j\sum_k r_{jk}}$$

Zは合成変量 Y に他ならず，グループ重心と原点の距離（$\sqrt{1-D^{(r)}}$）は，

$$\sqrt{1-D^{(r)}}=\sum_j r(Y,X_j)/n=\sigma_Y$$

となる。これが大きいほど Y の代表性が高い。

以上の性質は，多くの下位テスト項目の得点をまとめて合成得点を作成する場合に利用できる。

重みを1として合成変量を作成し，それと個別項目との関連を調べる方法としては，従来セントロイド法第1因子を使う例が多い。また項目をいくつかのグループに分類してそれぞれについて合成変量を求め，かつそれらと全項目との関係を調べる方法として，群セントロイド法（Group centroid method）の斜交解がある。

本稿で述べる方法は，群セントロイド法のように事前に分類を与えるのではなく，合成変量の数（項目グループの数）のみを指定して，代表性の高い（$D^{(r)}$ が小さい）合成変量を内成的に作り出す方法であるといえる。より適した分類構造ベクトルを用いるのであるから，項目分類と合せて群セントロイド法の計算を行えば，はじめから恣意的に分類を与える通常の場合よりもよい単純構造化を果せると期待できる。

4. 系統的クラスター化

系統的クラスター化（Hierarchical clustering）[1]とは下位のグループが上位のグループに順次包含されていく形をいう。

まず，項目の数だけグループが存在する状態から出発する。はじめに，Dが最小となる2項目を同じグループとする。次に再びグループ（所属数1個のときは，各項目）を二つずつ組合せて D が最小となるものを合併する，……という手順をふんで所定のグループ数に達するまで行なう。したがって，一度作られたグループは分解されることがなく，それより大きいグループの中に組込まれることになる（図2）。

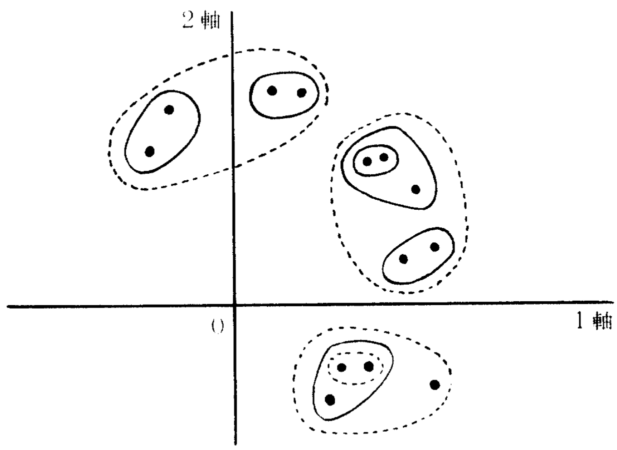

図2　系統的逐次合併

このような系統的クラスター化は，クラスターの数を固定したとき，そこで用いるクラスターの測度定義に関して最適であることを保証しない。この方式は実用の立場との‘妥協’である。項目のあらゆる可能な組合せを扱わずに少い計算量で実用上満足し得るグルーピングを達成することをねらっている。ここで扱う項目分類でもその立場をとる。目的によっては，全項目がこのように系統的に分類される方がより望ましい場合も多いであろう。

5. 分類の計算手順

ここでいう系統的項目分類は，二つの項目グループA，B を合併して新しいグループ C を作ったとき，測度 D_C が最小となる組合せA，B を全体の中から探すという手順を反復して行なう。

二つのグループA，Bを合併しグループ C を作ったときの測度 D_C は次のようになる。

グループ A，B の項目の数を n_A，n_B，グループ C の項目数を n_C とすると，

$$n_C=n_A+n_B$$

n_C 個の項目に関する内積行列を $H=\{h_{jk}\}$ とする。

$$h_{jk}=\sum_i x_{ij}x_{ik}/N$$

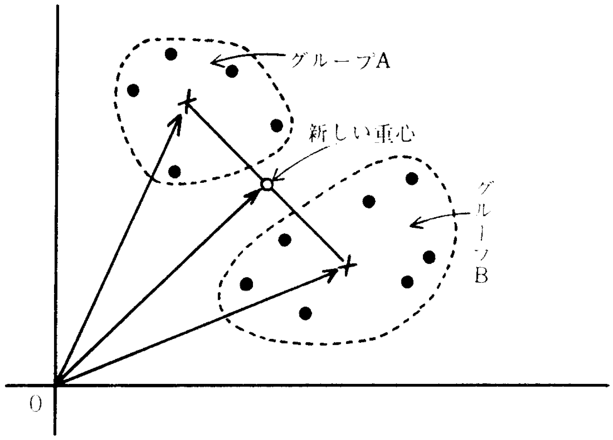

図3　グループ合併による重心の移動

行列 $\mathbf{H}_C$ をグループ A と B の項目ごとにまとめて次のように考えておく。

$$\mathbf{H}_C=\begin{bmatrix}\mathbf{H}_A & \mathbf{H}_{AB}\\ \mathbf{H}_{BA} & \mathbf{H}_B\end{bmatrix}$$

$\mathbf{H}_A$ は n_A 次, $\mathbf{H}_B$ は n_B 次の正方行列だが, $\mathbf{H}_{AB}$, $\mathbf{H}_{BA}$ は一般には矩形行列である。$\mathbf{D}_A$, $\mathbf{D}_B$ をそれぞれ $\mathbf{H}_A$, $\mathbf{H}_B$ の対角要素のみからなる対角行列とする。

新しい集中度 D_C はベクトル $\mathbf{1}$ を n_A 次または n_B 次の要素 1 のベクトルとして,

$$D_C=\frac{\mathbf{1}'\mathbf{D}_A\mathbf{1}+\mathbf{1}'\mathbf{D}_B\mathbf{1}}{n_C}-\frac{\mathbf{1}'\mathbf{H}_A\mathbf{1}+\mathbf{1}'\mathbf{H}_B\mathbf{1}+2\mathbf{1}'\mathbf{H}_{AB}\mathbf{1}}{n_C{}^2}$$

旧集中度 D_A, D_B を代入すれば,

$$n_CD_C=n_AD_A+n_BD_B+\left\{\frac{\mathbf{1}'\mathbf{H}_A\mathbf{1}}{n_A}+\frac{\mathbf{1}'\mathbf{H}_B\mathbf{1}}{n_B}-\frac{\mathbf{1}'\mathbf{H}_A\mathbf{1}+\mathbf{1}'\mathbf{H}_B\mathbf{1}+2\,\mathbf{1}'\mathbf{H}_{AB}\mathbf{1}}{n_C}\right\}$$

いま, 3で示したようなグループA, Bの項目合成得点（単純加算平均）を Y_A, Y_B とする。それらの平均, 分散を $\bar{Y}_A$, $\bar{Y}_B$, $\sigma^2(Y_A)$, $\sigma^2(Y_B)$, 相関を $r(Y_A, Y_B)$ とすると

$$n_CD_C=n_AD_A+n_BD_B+\frac{n_An_B}{n_C}\Big(\sigma^2(Y_A)+\sigma^2(Y_B)-2r(Y_A,Y_B)\sigma(Y_A)\sigma(Y_B)\Big)+\frac{n_An_B}{n_C}(\bar{Y}_A^2+\bar{Y}_B^2-2\bar{Y}_A\bar{Y}_B)$$

これから, わかるように新しい集中度 D_C は, 合併する二つのグループの集中性が高いと同時に, それらの重心間距離が近いとき値が小さくなる。

全項目が平均0, 分散 1 に基準化されていれば, w_A, w_Bを項目数に比例するウエイトとして,

$$w_A=n_A/(n_A+n_B),\quad w_B=n_B/(n_A+n_B)$$

$$1-D_C^{(r)}=\sigma^2(Y_C)=w_A{}^2\sigma^2(Y_A)+w_B{}^2\sigma^2(Y_B)+2w_Aw_B\,r(Y_A,Y_B)\,\sigma(Y_A)\,\sigma(Y_B)$$

のように表わせる。概していえば, $D^{(r)}$ が小さいことは $r(Y_A, Y_B)$ が大きいことと対応する。しかし, Dが最小となる組を求めて合併するという操作は合成変量間の相関最大のものを合併する方式とは異る。

各項目を分類するに当り, 符号反転を行うのが妥当な場合がある。項目（または変数）の数値の正, 負の方向が便宜的に与えられていることが, 実際場面では多い。

たとえば, 分類されるべき項目が S.D 法 (Semantic Differential method) におけるスケールのようなときで, 「明るい—暗い」において「明るい」をプラス側にとっても, 「暗い」をプラス側にとっても本質的な差異はない。図4において項目 j と k はグループ m′ として

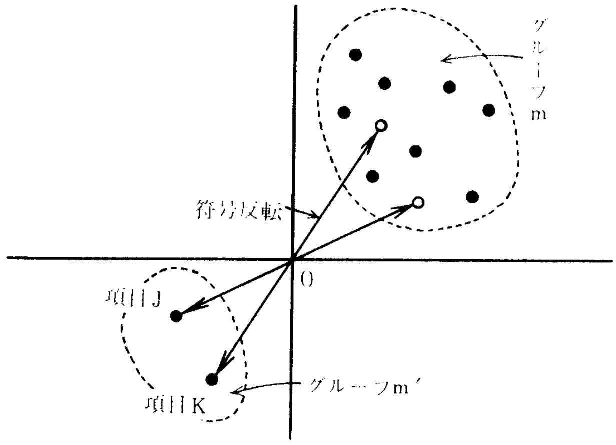

図4　符号反転

1グループを構成しているが, これらの項目の符号を反転すれば, 当然グループ m に含まれるべきであろう。グループの合併はこのような符号反転の操作を伴いつつ行なわなければならない。

以下任意の項目間内積行列 $\mathbf{H}$ について符号反転を行いつつ分類を進める計算を述べる（ここでいう内積行列は, 積和平均, 共分散, あるいは相関のいずれでもよい）。

まず, 各グループ ℓ に, どの項目が所属し, どの項目が所属しないかを指定するベクトル $\mathbf{v}_\ell$ が対応するとする。この $\mathbf{v}_\ell$ を分類構造ベクトルと呼ぶことにする。$\mathbf{v}_\ell$ の要素を v_j とすると,

$$\mathbf{v}_\ell'=\{v_1, v_2, \cdots\cdots\cdots\cdots, v_j, \cdots\cdots\cdots\cdots v_n\}$$

ここで n は分類対象の項目総数。j の順序は与えられた内積行列における項目の配列と同じである。

要素 v_j の値を次の通りとする。

$$v_j\begin{cases}=1\cdots\cdots\text{項目 j がグループmに属するときで, 符号反転の必要がないとき。}\\=-1\cdots\text{項目 j がグループmに属するが, 符号反転する必要があるとき。}\\=0\cdots\cdots\text{項目 j がグループmに属さないとき。}\end{cases}$$

全体の内積行列を**H**として，グループℓの集中度D_ℓは，

$$D_\ell = \mathbf{v}_\ell' \mathbf{D}_H \mathbf{v}_\ell / n_\ell - \mathbf{v}_\ell' \mathbf{H} \mathbf{v}_\ell / n_\ell^2$$

グループA，B を合併して C を作る場合でいえば，

$$D_C = (\mathbf{v}_A+\mathbf{v}_B)' \mathbf{D}_H (\mathbf{v}_A+\mathbf{v}_B) / (n_A+n_B) - (\mathbf{v}_A+\mathbf{v}_B)' \mathbf{H} (\mathbf{v}_A+\mathbf{v}_B) / (n_A+n_B)^2$$

となる。この D_C が最小となる項目の，組合せを探すわけであるが，上述のように **H** が相関行列 **R** であれば，

$$D_C^{(r)} = 1 - (\mathbf{v}_A+\mathbf{v}_B)' \mathbf{R} (\mathbf{v}_A+\mathbf{v}_B) / (n_A+n_B)^2$$

のように簡単になり，$D_C^{(r)}$ の最小を求める代わりに上式の右辺第二項の最大値を求めても同じである。

合併がある程度進んだ段階で D_C 最小となる組合せを求めるに当って，各組合せについてグループ全体を符号反転する場合としない場合の二通りを計算し，そのうちより小さい方をその組合せの D_C の値とする。それを行なうには $\mathbf{v}_A^*$ を $\mathbf{v}_A$ の反転ベクトル

$$\mathbf{v}_A^* = -\mathbf{v}_A$$

として，上式 $\mathbf{v}_A$ の値として，$\mathbf{v}_A$ および $\mathbf{v}_A^*$ の二通りの値を入れて計算する（$\mathbf{v}_B$はそのまま）。

この処理を全組合せについて実行し，最小となる組合せが得られたら，新しい $\mathbf{v}_C$ を，$\mathbf{v}_A$ および $\mathbf{v}_A^*$ のどちらが D_C 最小に効いたかに応じて

$$\mathbf{v}_C = (\mathbf{v}_A+\mathbf{v}_B) \quad \text{または} \quad \mathbf{v}_C = (\mathbf{v}_A^*+\mathbf{v}_B)$$

とする。また新グループの項目数 n_C を

$$n_C = n_A+n_B$$

とすればよい。

各項目が個々グループであるような最も細い分類状態からスタートするから，符号反転処理によって，ひとたび同じグループに編入されたら次に符号を反転するときは，グループ全体と一緒に反転することになる。

具体的な計算手順は次のようになる。実際の計算プログラムでは，計算量を小さくするための細部の工夫を加える必要がある。

① 諸データの入力，および前準備。

(1) m…項目の数

(2) 内積行列 **H** を分解して，$\mathbf{D}_H$ と $\mathbf{H}_0$ を作る。

$$D_H = \begin{pmatrix} h_{11} & & & 0 \\ & h_{22} & & \\ & & \ddots & \\ 0 & & & h_{mm} \end{pmatrix} \qquad H_0 = \begin{pmatrix} 0 & & h_{jk} \\ & \ddots & \\ h_{kj} & & 0 \end{pmatrix}$$

(3) 初期分類構造行列 $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m\}$ を作る。

$$V = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}$$

(4) グループ項目数を表わすベクトル $\mathbf{n} = [n_1, n_2, \cdots, n_m]$ を作る。初期値は，

$$\mathbf{n} = [1, 1, \cdots, 1]$$

② 集中度 D の計算

(1) $\mathbf{T} = \mathbf{V}' \mathbf{D}_H \mathbf{V}$ および $\mathbf{S} = \mathbf{V}' \mathbf{H}_0 \mathbf{V}$ の計算

$$T = \begin{pmatrix} t_{11} & & & 0 \\ & t_{22} & & \\ & & \ddots & \\ 0 & & & t_{mm} \end{pmatrix} \qquad S = \begin{pmatrix} \ddots & & S_{jk} \\ & \ddots & \\ S_{kj} & & \ddots \end{pmatrix}$$

(2) S の対角線下には，$-s_{jk}$ の値を入れる。

$$S^* = \begin{pmatrix} \ddots & & S^*_{jk}(=S_{jk}) \\ & \ddots & \\ S^*_{jk}(=-S_{jk}) & & \ddots \end{pmatrix}$$

(3) 集中度 D_{jk} を j と k ($j \neq k$) の全要素について計算する。

$$D_{jk} = \frac{(n_j+n_k-1)(t_{jj}+t_{kk}) - (s^*_{jj}+s^*_{kk}+2s^*_{jk})}{(n_j+n_k)^2}$$

(4) $\mathrm{Min}\{D_{jk}\}$ となる j と k を求める。

③ グループの合併

(1) $\mathrm{Min}\{D_{jk}\}$ となる j と k において $\mathbf{v}_j$ を $\mathbf{v}_k$ に合併する。$\mathbf{v}_j$ は 0 ベクトルにする。

j＞k ならば符号反転する。

$$V = \begin{array}{c} \begin{array}{ccccc} & & k & j & \end{array} \\ \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & -1 & 0 & \\ & & & & 1 \end{bmatrix} \end{array}$$

j＜k ならばそのまま合併。

$$V = \begin{array}{c} \begin{array}{ccccc} & & j & k & \end{array} \\ \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 0 & 1 & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} \end{array}$$

(2) n_j を n_k に合併する。n_j は 0 とする。

$$\begin{array}{c} \begin{array}{ccccc} & k & & j & \end{array} \\ n = [\cdots\cdots & 1 & \cdots\cdots & 0 & \cdots\cdots] \end{array}$$

④ 以下②，③の手順を反復する。分類構造行列**V**において 0 ベクトルが増して所定のグループ数に達したら終了し，**V**，**n** を出力する。

入力行列の中から特定の項目を除いて行なうときは，分類構造行列 **V** の対応する列を 0 ベクトルとして開始すればよい。また，分類を行なうとき特定の項目どうし

は同一グループでなくては困る場合がある。このときも同様に V の初期値をそのように指定しておけばよい。

6. 計 算 例

次に具体的な数値例を挙げる。例示のデータは，いわゆる‘三島由紀夫’事件の直後に工科系男子大学生約 40 名より得た意見調査の結果である*)。調査項目および被調査者ともに少数であるため，資料的価値は乏しいものである。ここでは計算の例示用としてのみ使うことにする。

データは，事件に対する 17 の意見あるいは感想の一つ一つについて「まったく反対」から「まったく賛成」まで 5 段階に評定してもらったもので，「まったく賛成」が 5 点，「まったく反対」が 1 点となる 5 点スケール値に直して用いた。

項目間相関行列を入力し（全項目について平均 0，分散 1 に基準化して扱う），3 グループに分類したあと，それによる群セントロイド法（斜交解）の結果を表 1 にまとめて示す。表 1 では 17 の項目が 5 グループにまとまったところから，最終的に 3 グループになるまでの合併過程が分るようにした。3 グループになったところで，グループⅠ，Ⅱ，Ⅲ別に項目合成得点を求め，それら相互間および個別項目との相関係数を掲げてある。（群セントロイド法の計算法による）* を付した項目はそれぞれのグループの中で合成得点を作る際，符号をマ

注 *）名古屋大学教育心理学教室の中島史明氏の調査による。入力相関行列は表 2 に示す通りである。

表 1　　系統的項目分類の計算例

‘三島’事件に関する意見項目	合併の順序 5グループ	合併の順序 4グループ	合併の順序 3グループ	合成得点と個別項目値との相関 Ⅰ	Ⅱ	Ⅲ
1）民主的秩序の破壊者である。	Ⅰ (0.633)	Ⅰ (0.633)	Ⅰ (0.633)	0.812	−0.382	0.110
3）彼らをこんな行動に走らす危機感がいまの日本にそれほどあるとは思われない。				0.683	−0.154	0.050
6）この事件は非常に時代ばなれした感じがする。				0.784	−0.374	0.301
2）現代の‘いらだしさ’に正面から立ちむかった彼らの行動力には大いに共感できる。	Ⅱ′ (0.851)	Ⅱ (0.879)	Ⅱ (0.879)	−0.197	0.762	−0.489
4）思想と行動とを合致させようという生き方には賛成である。				−0.209	0.794	−0.312
*5）彼等の行動はまったく理解できない。				0.489	−0.668	0.141
7）安逸・惰性に流れる社会現象に追随する政治の姿勢へのきびしい抵抗である。				−0.061	0.773	0.278
10）何もせぬ文化人より死をかけて行動した点は立派である。				−0.292	0.737	0.537
14）われわれの手で日本を守るのだという信念には心をうたれる。				0.301	0.625	0.047
*9）正常な精神状態から発した行動とは認めがたい。	Ⅱ″ (0.733)			0.267	−0.625	−0.021
11）これは五・一五事件，二・二六事件などのテロ事件と相通じるものだ。				−0.283	0.617	−0.434
*12）まったく馬鹿げている。				0.463	−0.818	0.315
8）これが呼び水になってこうした事件が続発するのがこわい。	Ⅲ′ (0.600)	Ⅲ′ (0.600)	Ⅲ (0.688)	−0.044	−0.333	0.648
15）彼がそんなことをするとは信じられない。				−0.009	0.143	0.537
16）こうした事件が外国に悪い印象を与えるだろうと心配する。				0.307	−0.339	0.779
*13）切腹はやっぱりどこかカッコよいものだ。	Ⅲ″ (0.673)	Ⅲ″ (0.673)		−0.147	0.415	−0.626
*17）芸術至上主義者の行動である。				−0.272	0.322	−0.742
合成得点間の相関			Ⅰ	1.000	−0.399	0.202
			Ⅱ	−0.399	1.000	−0.380
			Ⅲ	0.202	−0.380	1.000

（注）「合併順序」の中の（ ）の数値は α 係数。

イナスとして扱うべきことを示している。また各下位グループにおける（　）内の数値は，それぞれのグループの中の項目群を同一内容の測定のくりかえしとみなした場合における信頼性の測度—Cronbach の α 係数である。α 係数は必ずしも高いといえないが，強いて全項目を三つの下位尺度に分割しようとするならば，表1のⅠ，Ⅱ，Ⅲの別に行なうのがよいことを示している。α 係数の高いグループ Ⅱ の項目群に主眼を置くならば，Ⅰ，Ⅲの項目は除外すべきことは，項目の意味内容からみても首肯できることと思う。

なお合成得点の算出は，内積行列 H を相関行列として行なっているので，項目ごとの粗点から合成するときは，重みは1でなく，標準偏差 σ で除さなければならない。

項目の分類が終了したとき，それらの全体像を図上に表現したい場合がある。自動分類自身は，グラフ化に必要な個々の項目の座標値を与えない。そのときは別に因子抽出を行なう必要がある。上の群セントロイド法（斜交解）もその1つであるが，因子負荷量を求める計算はなにを用いてもよいだろう。グルーピングが完了しているという状態を活かし，なるべく同じグループに属する項目が近寄り，グループ重心が離れるような方法を適用するのが望ましい。群セントロイド法(斜交解）では‘斜交’という特徴が図示には不便である。ここでは，同じく群セントロイド法の直交解でそれを行なってみた（図5）。比較のため，通常のセントロイド解—バリマックス回転の結果を合せて掲げる。図5(1)で＊の項目は符号反転したあとの位置を示す。バリマックス回転後の図5(2)でもそれらの項目は反転してある。この計算例では両者の違いは小さい。しかし一般には強制的にグループを定めて軸を導くのでバリマックス回転の場合より図的まとまりはよいといえそうである。

なお，本稿では相関行列に関して「分類」を行なう場合，対角要素を1——すなわち共通性を1として扱った。この計算例についても同様であった。これを共通因子空間に限定して行なうことは可能である。

項目 j のベクトルの長さを h_j（$\leqq 1$）として集中度 $D^{(r)}$ を，

$$D^{(r)}=\sum_j h_j{}^2/n-\sum_j\sum_k r_{jk}/n^2$$

と考えればよく，計算上異る問題はない。しかし，共通性自身をラフな推定法によって得た場合，相関行列の非負定符号性（positive semi-definite）を損う危険があり好ましくない。$h_j{}^2=1$ として項目分類を行なえば，本来共通性の小さい項目は孤立グループとして合併から取残されてくる筈である。

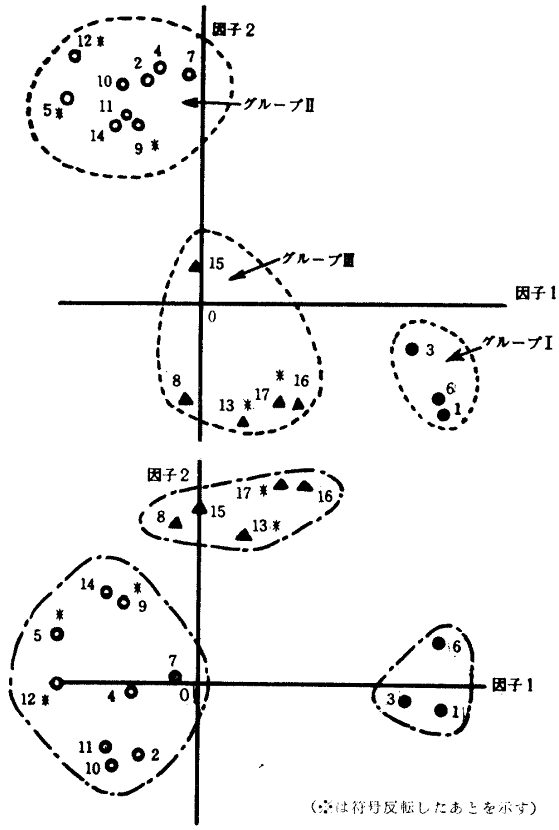

図5(1)　群セントロイド法（直交解）による

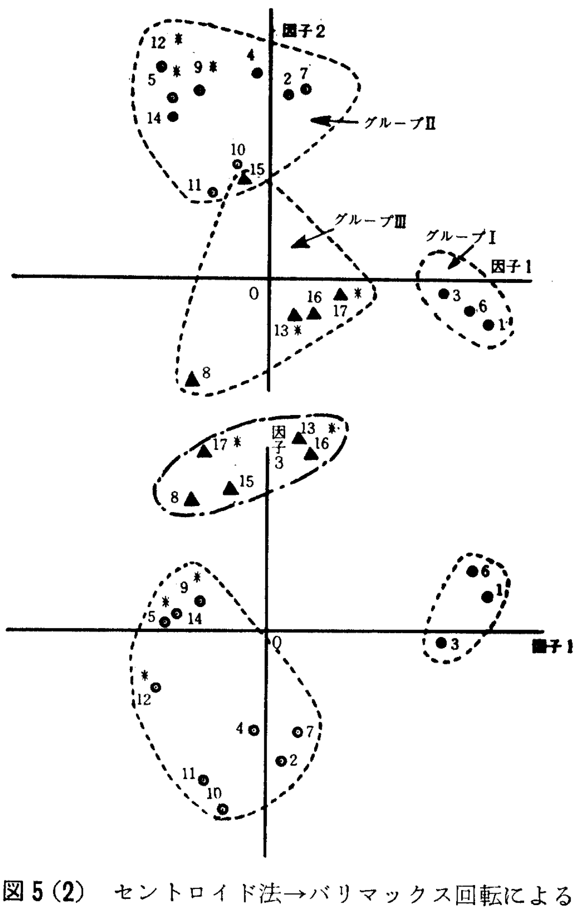

図5(2)　セントロイド法→バリマックス回転による

表 2　　入力相関行列

項目	1	3	6	2	4	5	7	10	14	9	11	12	8	15	16	13	17
1	100	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3	31	100	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
6	54	25	100	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2	−15	−08	−22	100	—	—	—	—	—	—	—	—	—	—	—	—	—
4	−27	−09	−12	66	100	—	—	—	—	—	—	—	—	—	—	—	—
5	35	35	41	−47	−51	100	—	—	—	—	—	—	—	—	—	—	—
7	−07	09	−16	59	54	−36	100	—	—	—	—	—	—	—	—	—	—
10	−25	−14	−28	67	58	−30	61	100	—	—	—	—	—	—	—	—	—
14	−33	05	−40	36	37	−47	44	38	100	—	—	—	—	—	—	—	—
9	24	29	08	−28	−42	39	−44	−30	−24	100	—	—	—	—	—	—	—
11	−37	01	−29	34	40	−19	51	43	29	−31	100	—	—	—	—	—	—
12	41	19	45	−53	−61	60	−47	−45	−47	64	−49	100	—	—	—	—	—
8	−13	03	−00	−47	−28	17	−31	−24	13	20	−31	28	100	—	—	—	—
15	−16	−03	17	−09	22	−03	01	−17	42	−37	−05	−17	18	100	—	—	—
16	31	−00	40	−34	−37	19	−16	−31	−03	−01	−40	39	53	29	100	—	—
13	−18	05	−21	40	36	−05	33	59	23	−01	39	−30	−16	−10	−32	100	—
17	−17	−22	−23	33	25	−10	14	49	13	−10	29	−25	−29	−22	−46	51	100

（小数点は省略。項目の配列順は表1に合せてある。）

7. おわりに

最後に，ここで述べた項目分類について一，二の問題点を加えておこう。

第一に，結果の安定性の問題がある。系統的クラスター化の方法はすでにできている下位グループを順次合併していくという方式であった。このことは次のことを危惧させる。すなわち，はじめの項目のセット次第で分類の結果が大きく左右されないか，例えば計算例の17項目のうち一，二の項目を欠いた場合に内容的に同等な解が得られるかどうかの問題である。

現段階で結果の安定性を一般的に保証することはむずかしい。実際的には試みに若干の項目を除外して数回計算を反復し，分類結果の安定性をチェックするのがよいと思われる。

次に，グループ合併の打切り基準の問題がある。グループ数をはじめに指定できないとき合併の進行を打切る方法としては，あらかじめ集中度 D の下限値を指定してグループの一つがそれを上廻る直前の段階で終了とする方法が一つ考えられる。また，Cronbach の α 係数などを利用して基準をもうける方法もよいであろう。

これらの点については，今後なお検討する必要がある。

引用文献

1. Johnson, S.C.　Hierarchical clustering schemes. Psychometrika, 1967, 241—254.
2. 京極純一　n元度数分布の数量化について．統数研彙報，1967，15，140—160.